School-Wide Screening

Hugh W. Catts, Ph.D.
University of Kansas

National SEA Conference on SLD Determination
Kansas City, MO
April 19-21, 2006

Early Identification
Early Identification

Early Intervention

Reduced Risk for RD

---

Early Intervention

- Critical because children who start out as poor readers generally continue to be poor readers
- Poor reading achievement quickly leads to a host of negative consequences
Negative Consequences

- Low motivation
- Negative expectations
- Limited practice
- Academic failure
Negative Consequences

- Low motivation
- Negative expectations
- Limited practice
- Academic failure

School-Wide Screening

- Importance of accuracy
- What to measure
- Current screening tools
- New directions in research
- Conclusions
Screening

- Screening tests have a long tradition in health professions.
- Used to detect potential health problems in an individual who doesn’t show symptoms.
- Once identified, follow-up testing is conducted, and if required, intervention is initiated to prevent or limit the condition or disease.
- Common screening tests include tests for high cholesterol, early signs of cancer, depression, or hearing problems.

Screening in Schools

- Screening tests also have a long tradition in education.
- Typically administered in kindergarten or first grade with the purpose of identifying children at risk for academic problems.
- Screening takes on a more prominent role in a RTI framework.
Screening Accuracy

- Particular attention is given to the accuracy of screening instruments
- Errors in identification can be costly
  - over identification
  - under identification

Public Health

- Over identification
  - expense of additional testing
  - unnecessary worry
- Under identification
  - miss serious health problem
Education

- Over identification
  - expense of additional testing
  - expense of early intervention services
- Under identification
  - miss opportunity for early intervention

Clinical Decision Making Model
Clinical Decision Making Model

Screen

<table>
<thead>
<tr>
<th></th>
<th>At risk</th>
<th>Not at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Positive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False Negative</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Outcome

Normal

RD

<table>
<thead>
<tr>
<th></th>
<th>True Positive</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

True Negative

| d         |               |

Total % Correct

\( \frac{a+d}{\text{total}} \)
### Screening Accuracy

<table>
<thead>
<tr>
<th>Screen</th>
<th>At risk</th>
<th>Not at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>Normal</td>
<td>10</td>
<td>370</td>
</tr>
</tbody>
</table>

Base rate 5%

Total % Correct 94%

### Clinical Decision Making Model

<table>
<thead>
<tr>
<th>Screen</th>
<th>At risk</th>
<th>Not at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome</td>
<td>True Positive</td>
<td>False Negative</td>
</tr>
<tr>
<td>Normal</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>False Positive</td>
<td>c</td>
<td>True Negative d</td>
</tr>
</tbody>
</table>

Sensitivity: \( \frac{a}{a + b} \)

Specificity: \( \frac{d}{c + d} \)
Clinical Decision Making Model

<table>
<thead>
<tr>
<th>Screen</th>
<th>At risk</th>
<th>Not at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Positive</td>
<td>a</td>
<td>False Negative</td>
</tr>
<tr>
<td>False Positive</td>
<td>c</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

Sensitivity: \( \frac{a}{a + b} \)
Specificity: \( \frac{d}{c + d} \)

Accuracy of Screening is determined by ...

- How well your instrument separates those who eventually will have a problem from those who will not
- What you choose as a cut-off score
The Ultimate Screen

Good Readers | Poor Readers
---|---
Number of errors

<table>
<thead>
<tr>
<th>TP</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FP</th>
<th>TN</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>
More Typical Screen

More Typical Screen

<table>
<thead>
<tr>
<th>TP</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FP</th>
<th>TN</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>80</td>
</tr>
</tbody>
</table>
More Typical Screen

<table>
<thead>
<tr>
<th>TP</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>95</td>
<td>5</td>
</tr>
<tr>
<td>FP</td>
<td>TN</td>
</tr>
<tr>
<td>35</td>
<td>65</td>
</tr>
</tbody>
</table>

ROC Curve

http://www.anaesthetist.com/mnm/stats/roc/
Base Rate

- The proportion of population that has the condition
- Sometimes base rate is straightforward
  - children with severe/multiple handicaps
- Base rate of RD is not clear-cut
- Reading achievement is continuously distributed with no clear demarcation between good and poor readers
- Doesn’t follow a categorical model
CATEGORICAL MODEL

"disorder" (KIDS WHO HAVE "IT")

"normal" (KIDS WHO DON'T HAVE "IT")

DIMENSIONAL MODEL

arbitrary cut-off

"disorder"  "normal"

# of KIDS WHO GOT THAT SCORE

SCORES on a MEASURE of LANGUAGE or READING

lowest score  average  highest score

HSS '91
Base Rate

- Determined in part by perspective on the purpose of early identification
  - Traditional LD model
    - base rate 5%
    - at-risk rate 15-20%
  - Prevention-oriented general ed model
    - base rate 20-30% at-risk rate 50%
- Dependent on resources

What to Measure?

- What is the criterion? What are we predicting to?
- Reading comprehension
- Reading comprehension involves a mixture of complex abilities
- Role of each changes over time
Predicting Comprehension

- word reading
  - letter knowledge
  - phonological awareness
- oral reading fluency
- vocabulary and grammar
- listening comprehension
Measures

- Need to be matched to abilities of children
- Should be consistent with the expectations of the curriculum
- Estimate of risk is a "moving target"

Measures

- Need to use multiple measures
- Most early predictors are only moderately correlated with reading
- Need a combination to attain high classification accuracy
- Measure more than once
Screening Tools

- Readily available
- Standardized
- Easily administered
- Accurate

Phonological Awareness Literacy Screening (PALS-K; Invernizzi, Juel, Swank, & Meier)

- [http://pals.virginia.edu](http://pals.virginia.edu)
- Measures kindergarten students’ literacy development with the following subtests:
  - Rhyme Awareness (group then individual if needed)
  - Beginning Sound Awareness (group then individual if needed)
  - Alphabet Knowledge
  - Letter Sounds
  - Spelling (group then individual if needed)
  - Concept of Word
- Takes approximately 30-45 minutes to complete
- A summed score is obtained which can be used to compare to benchmarks (fall and spring)
- PALS-PreK and PALS 1-3 also available
- Classification accuracy of combined PALS K, 2-3 to state assessment was 82%
Texas Primary Reading Inventory
(Foorman et al., 1998- www,tpri.org)

- Designed to be used by teachers to identify children at risk for RD and to further evaluate their strengths and weaknesses in reading-related skills
- 5 screens for K-2nd grade
- Designed to hold false negatives at a minimum
- Includes an inventory of secondary measures to help rule out false positives and inform instruction

TPRI (1998)
K (Dec) predicting end of 1st
Screen (shorten version)

- Letter-sound identification (10 items)
- Phoneme blending (8 items)
**TPRI (1998)**  
**K (Dec) predicting end of 1st**

<table>
<thead>
<tr>
<th>Screen (shorten version)</th>
<th>At risk</th>
<th>Not at risk</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD</td>
<td>Normal</td>
<td>Positive Predictive Power</td>
<td>Negative Predictive Power</td>
<td></td>
</tr>
<tr>
<td>RD</td>
<td>Normal</td>
<td>92</td>
<td>5</td>
<td>94.8%</td>
</tr>
<tr>
<td>RD</td>
<td>Normal</td>
<td>143</td>
<td>181</td>
<td>Base rate 23%</td>
</tr>
<tr>
<td>RD</td>
<td>Normal</td>
<td>39.1%</td>
<td>97.3%</td>
<td></td>
</tr>
</tbody>
</table>
Texas Primary Reading Inventory
(Foorman et al., 1998- www.tpri.org)

- Inventory of secondary measures (12 measures)
  - book and print awareness
  - rhyming
  - blending word parts
  - blending phonemes
  - deleting initial sounds
  - deleting final sounds
  - letter-name identification
  - letter to sound linking A & B
  - listening comprehension 1-3
- Most have 5 items
- Designed to progress for easy to difficult
- About 20 minutes to administer

Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

- Standardized and readily available
  - www.dibels.uoregon.edu
  - www.aimsweb.com
- Curriculum-Based Measurement Tool (CBM)
- Developed to monitor progress and inform instruction
CBM Tools

- Short assessments
- Most often involve speeded performance
- Multiple forms
- Tied to curriculum

CBM Tools

- Letter-Name Fluency
- Letter-Sound Fluency
- Initial-Sound Fluency
- Phoneme Segmentation Fluency
- Nonword Reading Fluency
- Word Identification Fluency
- Oral Reading Fluency
- Oral Retell Fluency
- Maze Fluency
CBM Tools

- Assessments given 3 or more times a year to evaluate growth in reading (meeting benchmarks)
- Each can be considered a screening opportunity

DIBELS

K (Fall) predicting end of 1st

Screen (Initial sound fluency, Letter name fluency)
DIBELS
K (Fall) predicting end of 1st*
Screen (Initial sound fluency, Letter name fluency)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>At risk</th>
<th>Not at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD</td>
<td>8577</td>
<td>1824</td>
</tr>
<tr>
<td>Normal</td>
<td>9345</td>
<td>12258</td>
</tr>
</tbody>
</table>

- Sensitivity: 82.5%
- Specificity: 56.7%
- Positive Predictive Power: 47.9%
- Negative Predictive Power: 87.0%
- Base rate: 32.5%
- Risk rate: 56.0%

*Adapted from summary data with important qualifications

First Grade Screening
TPRI (1998)
1st (Oct) predicting end of 1st

Screen (letter-sound, blending, word reading)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>At risk</th>
<th>Not at risk</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive Predictive Power</th>
<th>Negative Predictive Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD</td>
<td>111</td>
<td>8</td>
<td>93.3%</td>
<td>63.5%</td>
<td>38.8%</td>
<td>97.4%</td>
</tr>
<tr>
<td>Normal</td>
<td>175</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Base rate 19.9%
Risk rate 47.7%

DIBELS
1st NWF predicting end of 1st ORF
DIBELS

1st NWF predicting end of 1st ORF

<table>
<thead>
<tr>
<th>Outcome</th>
<th>At risk</th>
<th>Not at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>7477</td>
<td>2956</td>
</tr>
<tr>
<td>RD</td>
<td>5067</td>
<td>16544</td>
</tr>
</tbody>
</table>

- Sensitivity: 71.7%
- Specificity: 76.6%
- Base rate: 32.6%
- Risk rate: 39.1%
- Positive Predictive Power: 59.6%
- Negative Predictive Power: 86.4%

*Adapted from summary data with important qualifications

Dynamic Assessment

- Measurement of ability over time in order to monitor progress
- Measurement of learners' potential over the short term
- Assessor actively intervenes during the course of the assessment with the goal of intentionally inducing changes in the learner's current level of performance.
- "Mini-assessment" of response to intervention
O’Connor & Jenkins (1999)

Oct 1st predicting April 1st

Screen (phoneme seg, RLN, phoneme repetition)

<table>
<thead>
<tr>
<th></th>
<th>At risk</th>
<th>Not at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Base rate 5.1%

Sensitivity 100%
Specificity 87.3%

Risk rate 17.2%

Positive Predictive Power 29.7%
Negative Predictive Power 100%
O’Connor & Jenkins (1999)

- Dynamic Assessment
  - taught at-risk children phoneme segmentation using a set of test items
  - score based on the number of trials needed to master the task

O’Connor & Jenkins (1999)
Oct 1st predicting April 1st (dynamic)

<table>
<thead>
<tr>
<th></th>
<th>Screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>At risk</td>
<td>10</td>
</tr>
<tr>
<td>Not at risk</td>
<td>1</td>
</tr>
<tr>
<td>Normal</td>
<td>9</td>
</tr>
<tr>
<td>RD</td>
<td>195</td>
</tr>
</tbody>
</table>

- Sensitivity: 90.9%
- Specificity: 95.6%
- Positive Predictive Power: 52.6%
- Negative Predictive Power: 99.5%

- Base rate: 5.1%
- Risk rate: 8.8%
Compton, Fuchs, Fuchs, & Bryant (in press)

- Screened in 1st (Oct) predicting end of 2nd
- Measures
  - CTOPP Sound Matching
  - CTOPP Rapid Digit Naming
  - WJPB-R Oral Vocabulary
  - Word Identification Fluency (WIF)
    Initial level, 5-week slope

Grade 1
Word-Identification Fluency

Teacher: **Read**
**these words.**

Time: 1 minute.

- two
- for
- come
- because
- last
- from
- ...
Compton, Fuchs, Fuchs, & Bryant (in press)

- Screened in 1st (Oct) predicting end of 2nd
- Measures
  - CTOPP Sound Matching
  - CTOPP Rapid Digit Naming
  - WJPB-R Oral Vocabulary
  - Word Identification Fluency (WIF)
    Initial level, 5-week slope

---

Compton et al. (in press)

1st (Oct) predicting end of 2nd

Screen (includes WIF level & slope)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>RD</th>
<th>At risk</th>
<th>Not at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>35</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>155</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>94.6%</td>
<td>91.7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Positive Predictive Power</th>
<th>Negative Predictive Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>71.4%</td>
<td>98.7%</td>
</tr>
</tbody>
</table>
Beyond First grade

- Most common screening for Tier 2 has been measure of ORF
- ORF strongly correlated with 3rd grade state assessments
- High correlations do not necessarily translate into high sensitivity and specificity

<table>
<thead>
<tr>
<th>Measure</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSA (Good, Simmons, &amp; Kame'enui, 2001)</td>
<td>.73</td>
</tr>
<tr>
<td>FCAT-SSS (Buck &amp; Torgesen, 2003)</td>
<td>.70</td>
</tr>
<tr>
<td>ISAT (Sibley, Biwer, &amp; Hesch, 2001)</td>
<td>.79</td>
</tr>
<tr>
<td>ASA (Linner, 2001)</td>
<td>NA</td>
</tr>
<tr>
<td>CSAP (Shaw &amp; Shaw, 2002)</td>
<td>.80</td>
</tr>
<tr>
<td>MEAP (4th grade) (McGlinchey &amp; Hixson, 2004)</td>
<td>49–81</td>
</tr>
</tbody>
</table>
### Concurrent Validity

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive Predictive Power</th>
<th>Negative Predictive Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSA (Good, Simmons, &amp; Kame'enui, 2001)</td>
<td>.73</td>
<td>89.4</td>
<td>71.3</td>
<td>43.7</td>
<td>96.4</td>
</tr>
<tr>
<td>FCAT-SSS (Buck &amp; Torgesen, 2003)</td>
<td>.70</td>
<td>85.3</td>
<td>69.0</td>
<td>57.3</td>
<td>90.6</td>
</tr>
<tr>
<td>ISAT (Sibley, Biwer &amp; Hesch, 2001)</td>
<td>.79</td>
<td>93.8</td>
<td>74.5</td>
<td>37.5</td>
<td>98.6</td>
</tr>
<tr>
<td>ASA (Linner, 2001)</td>
<td>NA</td>
<td>89.7</td>
<td>74.3</td>
<td>44.3</td>
<td>96.9</td>
</tr>
<tr>
<td>CSAP (Shaw &amp; Shaw, 2002)</td>
<td>.80</td>
<td>80.0</td>
<td>62.8</td>
<td>42.9</td>
<td>90.0</td>
</tr>
<tr>
<td>MEAP (4th grade) (McGlinchey &amp; Hixson, 2004)</td>
<td>.49-.81</td>
<td>75.0</td>
<td>74.0</td>
<td>77.0</td>
<td>72.0</td>
</tr>
</tbody>
</table>

---

### CBM & State Assessments

- Reported results are usually much better
- Most reports only consider the low and high risk groups
- Students in the “some risk” category are not included
- Equally likely to have good vs. poor outcomes
- But results should be expected on the basis of the simple view
Possible Solutions

- Measurement of level and slope may help (e.g., dual discrepancy)
- Add assessments of language abilities
What have we learned about screening?

- Can identify children at risk for reading problems
- Can be done as early as the fall of kindergarten
- Need to choose measures carefully
- Must match measures to curriculum
  - letter naming
  - phonological awareness
  - word reading
  - text reading
- Must not forget about other factors related to comprehension
  - oral language

False positive rates are high and efforts need to be in place to limit the cost of over prediction

- Brief secondary assessments (TPRI)
- Duel discrepancy
- Short-term instruction (dynamic assessment)
- Tier 2 (RTI)